3Н

УДК 004.934.1'1

***А.В. Ниценко, В.Ю. Шелепов, Д.С. Азаренко***
Институт проблем искусственного интеллекта МОН Украины и НАН Украины,
г. Донецк, Украина
Государственный университет информатики и искусственного интеллекта,
г. Донецк, Украина
nav_box@mail.ru, shel@iai.dn.ua

# О компьютерном распознавании речи с учетом семантики произносимого

Для успешного распознавания речи человеком важнейшее значение имеет использование смысловой связи распознаваемых слов. Более того, успешность речевой коммуникации состоит, в конечном счете, в достижении совпадения смысла произносимого и воспринимаемого. Статья посвящена разработанной авторами программе распознавания речи, использующей смысловую сочетаемость слов в предложении. Изложение ведется на примере набора математических текстов.

## Введение

Все наши разработки по распознаванию речи до недавнего времени сводились к попыткам распознавать произносимые слова либо как цельные языковые объекты (распознавание по эталонам), либо на основании классификации более мелких языковых единиц (пофонемное распознавание). Распознавание носило, если можно так выразиться, механистический характер: каждое слово распознавалось отдельно от остальных на основе более или менее удачных алгоритмов и признаков классификации. В то же время, согласно точке зрения современной психолингвистики, человек, слушая речь другого, все время строит гипотезы относительно только что услышанного слова или даже только что услышанных фонем и при этом многократно мысленно перемещается от конца к началу услышанного и обратно, корректируя упомянутые гипотезы так, чтобы при этом слова выстраивались в некоторые смыслы в соответствии с контекстом произносимого. Эти смыслы и контекст возникают в уме слушающего как результат сопоставления того, что он услышал, с собственным жизненным опытом и знаниями об окружающем мире. Успешность коммуникации состоит, в конечном счете, в достижении совпадения смысла произносимого и воспринимаемого. Представляется, что без имитации такого механизма нельзя получить удовлетворительный результат компьютерного распознавания содержательной цепочки произносимых слов.

Конкретно в применении к нашим системам пофонемного распознавания это дает следующее. На многих этапах процесса распознавания возможен отказ от него. При распознавании между собой, например, какой-либо пары фонем это приводит к тому, что обе рассматриваемые фонемы следует включать в число допустимых. В итоге в общем случае результатом распознавания очередного произнесенного слова является не одно слово, а некоторый список слов – кандидатов на распознавание. Использование семантики призвано помочь выбрать слово из этого списка или хотя бы существенно сократить последний.

# Таблица сочетаемости

Условимся под термином «сочетаемость» двух слов понимать возможность следования одного из них непосредственно за другим в предложении. Очевидно, проявлением семантики является сочетаемость двух последовательно распознанных слов. Поэтому при распознавании очередного слова ряд слов-кандидатов можно будет отбросить, как не сочетающиеся с предыдущим.

Сочетаемость слов мы описываем с помощью таблицы (таблица сочетаемости), которую можно представить себе следующим образом. Слева и вверху ее находятся леммы (начальные формы) слов рабочего словаря. Если слово может идти следом за некоторым другим, то в соответствующей клетке таблицы проставляется символ «1», в противном случае эта клетка оставляется пустой. Леммы используются вместо словоформ ввиду необходимости сокращения размера таблицы, который является второй степенью от размера словаря.

Наилучшим способом автоматического заполнения таблицы является обработка и анализ реальных текстов. Обработка начинается с процедуры лемматизации – нахождения начальной формы каждого встретившегося слова. В процессе работы с текстом происходит пополнение как словаря для распознавания, так и таблицы сочетаемости. Для автоматической лемматизации используется программа лемматизатор, которая ставит в соответствие каждой словоформе соответствующую лемму. Быстрый поиск слов в словаре и таблице сочетаемости осуществляется за счет представления словников в виде дерева, так что глубина поиска оказывается равной длине слова.

Ввиду широко представленной в языке омонимии автоматическая лемматизация может приводить к неоднозначным результатам. Так при обработке текста 1-го тома книги Г.М. Фихтенгольца «Курс дифференциального и интегрального исчисления» лемматизатор предложил в качестве лемм слова «БЫЛИ» два варианта: «БЫЛЬ» и «БЫТЬ». Первый результат, очевидно, для математического текста является маловероятным. На самом деле леммой здесь является, конечно, глагол «БЫТЬ». Таким образом, подготовка текстов для пополнения таблицы сочетаемости требует участия человека. Поэтому является актуальным максимальное использование уже подготовленного текста.

# Пополнение таблицы сочетаемости

Начнем со следующих простых замечаний. Каждое слово в предложении может быть заменено синонимом, и это дает возможность пополнять таблицу, построенную с использованием заданного текста, используя словари синонимов. Союзу «И» может, очевидно, предшествовать и следовать за ним любое слово.

Далее, в отличие, скажем, от английского, простое русское предложение часто допускает почти произвольную перестановку слов. Например, предложение «Сестра быстро написала отцу письмо» допускает следующие варианты:

«Сестра написала отцу письмо быстро»,
«Сестра отцу письмо быстро написала»,
«Сестра письмо быстро написала отцу»,
«Быстро сестра письмо написала отцу»,
«Быстро письмо сестра написала отцу»

и так далее.

Каждый из этих вариантов может естественным образом звучать в устной речи со своим смысловым ударением.

Однако простое русское предложение может быть разбито на отрезки, называемые нами сегментами, в которых порядок слов является более жестким, но которые можно между собой переставлять.

### О членении простого русского предложения

1. Находим глагол или деепричастие. Выделяем их в отдельный сегмент. Если перед ними стоят наречия, включаем их в сегмент.

2. Выделяем в отдельный сегмент наречия, за которыми не следует глагол, причастие, деепричастие или прилагательное.

3. Выделяем в отдельный сегмент прилагательные, за которыми не следует существительное.

4. Выделяем в сегмент предлог, порядковое числительное, набор прилагательных и существительное в соответствующем падеже плюс цепочку существительных в родительном падеже, возможно с прилагательными или причастиями. Любые составляющие этой последовательности могут отсутствовать. Каждое существительное может быть заменено местоимением.

5. Выделяем в сегмент предлог, количественное числительное, набор прилагательных и существительное в соответствующем падеже плюс цепочку существительных в родительном падеже, возможно с прилагательными или причастиями. Любые составляющие этой последовательности могут отсутствовать. Каждое существительное может быть заменено местоимением.

При обсуждении проблемы синтаксического подчинения слов в предложении особенностью сегмента является то, что для каждого его слова, кроме «главного», подчиняющее слово находится в пределах этого же сегмента. В проблеме сочетаемости разбиение предложения на сегменты может быть использовано следующим образом.

Мы получим семантически и стилистически правильное предложение, если в пределах каждого сегмента удалим каждое прилагательное, причастие или порядковое числительное, за которым стоит существительное в той же форме. Мы получим семантически и стилистически правильное предложение, если в пределах каждого сегмента удалим каждое наречие, за которым стоит глагол. Тем самым мы получим новую допустимую последовательность слов и, соответственно, возможность для пополнения таблицы сочетаемости.

Далее, порядок слов внутри сегмента является достаточно жестким и произвольно менять его нежелательно. Сегменты же можно произвольным образом менять местами, что дает новые сочетания первых и последних слов соседних сегментов и возможность для пополнения таблицы сочетаемости.

## Система голосового набора математических текстов с учетом семантики произносимого

### Описание текстового редактора, работа с таблицей сочетаемости

Начнем с разработанной нами программы Voice Edit. Она написана с использованием существующего открытого кода программы Word Pad. Вот ее основное окно, в котором отображается набираемый словесный текст и формулы:

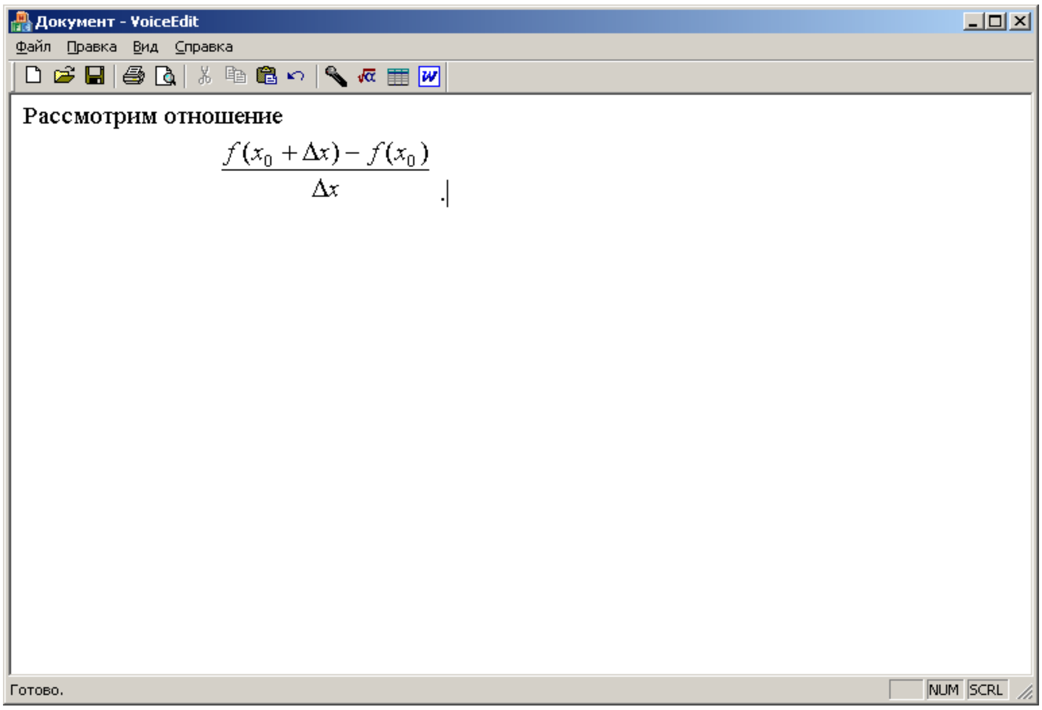



Рисунок 1 – Основное окно программы Voice Edit

Для набора формул используется программа Equation, которая может быть запущена нажатием кнопки со значком $\sqrt{\alpha}$ (о запуске Equation-а из программы для голосового ввода Voice Input см. ниже).

При нажатии кнопки с изображением микрофона загружается окно набора словесного текста:

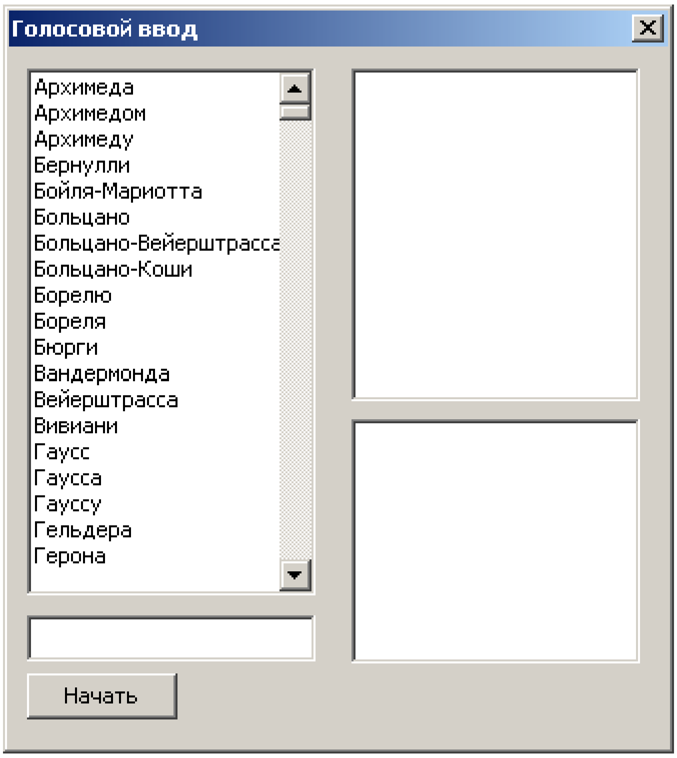



Рисунок 2 – Окно набора программы Voice Edit

Оно используется при голосовом вводе, но им можно пользоваться и для ввода текста с клавиатуры.

Словарь всех используемых словоформ, отображенный в левом списке, хранится в программе в виде дерева. Это позволяет в реальном времени искать в нем нужные слова. Представим себе, что мы находимся в каком-то месте набора текста. После того, как закончено предложение и поставлен знак его окончания (точка, восклицательный или вопросительный знак), первое слово нового предложения набирается в левом нижнем поле ввода. Его поиск в словаре автоматически начинается после нажатия пробела. Если слово найдено, оно вводится в текст в окне, изображенном на рис. 1. В противном случае подается звуковой сигнал о том, что слово не найдено (последнее возможно из-за ошибки при наборе и тогда пользователь вносит исправление и вторично нажимает пробел). При вводе второго слова предложения аналогичная процедура дополняется поиском леммы слова в таблице сочетаемости. Если слово найдено в словаре и сочетание его леммы с леммой предыдущего слова допускается таблицей сочетаемости, оно вводится в текст (рис. 1). Если слово найдено в словаре, но его лемма не допускается таблицей сочетаемости, то следует звуковой сигнал, отличный от упомянутого выше. При нажатии клавиши (+) на цифровой клавиатуре таблица пополняется нужным сочетанием лемм и слово попадает в текст (рис. 1).

Отметим, что использование звуковых сигналов позволяет осуществлять набор, глядя не на экран монитора, а на клавиатуру, что делает систему удобной для рядового пользователя, не искушенного в искусстве быстрого «набора в слепую».

## Голосовой ввод, взаимодействие с текстовым редактором

Для голосового ввода используется дополнительная программ Voice Input. Вот ее основное окно:

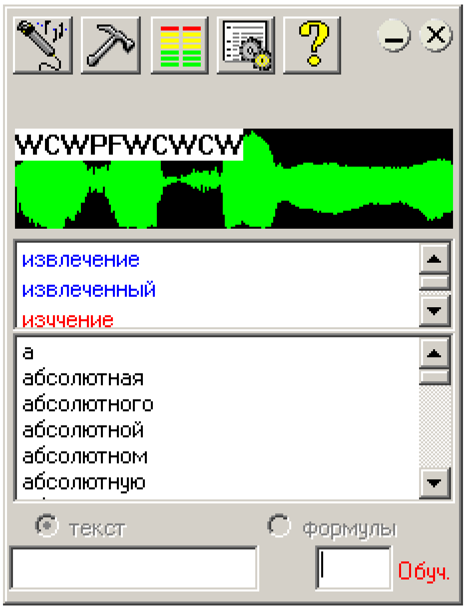



Рисунок 3 – Основное окно программы Voice Input

Она может работать в двух режимах: 1) голосовой ввод словесного текста; 2) голосовой ввод математических формул. В первом случае она взаимодействует с вышеописанной программой Voice Edit, во втором – с программой Equation.

Переход от первого режима ко второму осуществляется автоматически при загрузке программы Equation, которая запускается при русской раскладке клавиатуры нажатием сочетания клавиш «alt + ю» (формула будет набираться в той же строке окна – рис. 1) или «alt + .» (формула – с новой строки).

Программа голосового набора формул и работа с ней описана в работе [1]. Остановимся теперь на работе со словесным текстом. На рис. 3 представлено основное окно программы Voice Input в соответствующем режиме. В нижнем списке выведен перечень слов распознаваемого словаря. В качестве примера используем голосовой ввод фрагмента «При изучении хода изменения функции…». Произнесено слово «изучении». Строка WCWPFWCWCW – распознанная обобщенная транскрипция. Знак W обозначает гласную, C – согласную, P – паузу, F – шипящую. Сочетание PF в данном случае соответствует аффрикате *Ч*.

В верхнем списке мы видим часть перечня распознанных лемм. Полностью этот перечень в данном случае выглядит так:

закон, заполнить, заполнять, знакомый, значение, извлечение, извлеченный, изучение, изученный, облегчение, обобщение, обращение, отбрасывать, отброшенный, перегруппировать, передоказывать, перемещение, перенесение, переходить, подвешенный, подмеченный, полупериметр, получаемый, получение, полученный, помещенный, понятие, поочередно, предварительный, представлять, привлечение, приращение, приуроченный, противный, протяжение, проходимый, проходить, прохождение, рассуждать, рассуждение, расширение, расширенный, сгущение, характеризовать, характерный, эллипсоид, (1)

Он оказался большим потому, что здесь мы используем при распознавании лишь самую общую транскрипцию в рамках широкой фонетической классификации. Поскольку искомое слово должно следовать за уже распознанным словом «при», программа Voice Edit, используя таблицу сочетаемости, отобрала из этих лемм лишь шесть, отображенных в правом верхнем списке окна на рис. 4:

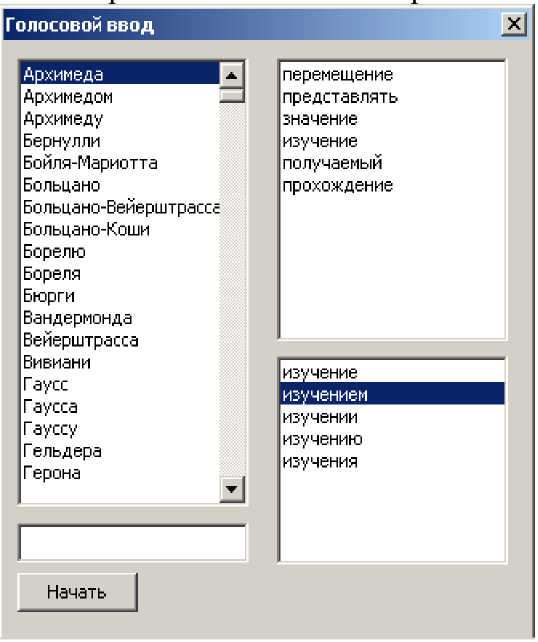



Рисунок 4 – Окно редактора программы Voice Edit при распознавании слова «изучении»

В этом коротком списке программа Voice Input с помощью эталона (о создании эталона см. в соответствующем разделе) распознала слово «изучение» и в правом нижнем списке вывела все его словоформы, допускаемые распознанной транскрипцией. Курсор автоматически перешел в этот список. Остается выбрать нужную словоформу «изучении». После нажатия «Enter» слово обрабатывается программой Voice Edit уже без использования таблицы сочетаемости и попадет в текст:

Если словоформа в правом нижнем списке на рис. 4 оказалась единственной, то она передается в текст автоматически, без вмешательства пользователя.

Появление вместо одной словоформы списка словоформ на сегодняшний день является неизбежным, поскольку большинство русских окончаний являются безударными, и их точное пофонемное распознавание на сегодня является практически невозможным. В связи с этим для выделения заранее нужной словоформы мы применяем следующий механизм. Перед произнесением слова можно, нажав предварительно пробел, ввести с клавиатуры последний символ или любое число заключительных символов словоформы, которая будет произноситься. Тогда логичным будет использование такого механизма и для сокращения числа кандидатов (рис. 4) путем указания с клавиатуры первого символа или любого числа начальных символов произносимой словоформы. Таким образом, программе можно «помогать» в распознавании, нажимая перед произнесением слова

первый символ (набор символов) – пробел – последний символ (набор символов)

или

первый символ (набор символов)

или

пробел – последний символ (набор символов).

Можно отказаться от использования клавиатуры и выбирать нужный вариант из предлагаемых списков с помощью мыши.

## О процедуре создания эталонов

Если список лемм (верхний список на рис. 4) содержит более одной леммы, то по указанию пользователя (щелчок мыши на элементе списка или выделение его курсором и нажтие “Enter”) по исходному звуковому файлу строится эталон и ему сопоставляется соответствующая лемма. Тогда, как показывает опыт, в подавляющем большинстве случаев при распознавании с использованием алгоритма DTW для любой словоформы этого слова построенный эталон окажется ближайшим и, следовательно, она будет ассоциироваться с указанной леммой. Исключения составляют ситуации типа «ИДТИ-ШЕЛ», «ЧЕЛОВЕК-ЛЮДИ». Далее, компьютер отбирает из всех распознанных словоформ список словоформ, отвечающих распознанному слову и соответствующих записанному сигналу. Это собственно и является результатом распознавания. При произнесении в дальнейшем других словоформ слова, для которого создан эталон, компьютер будет распознавать эти словоформы, используя имеющийся эталон. Подчеркнем, что эталон строится по произвольной произнесенной словоформе, получает имя соответствующей леммы, и по нему распознаются (с точностью до леммы) все другие словоформы этого слова. Если лемма распознанного слова (верхний список на рис. 4) является в этом списке единственной, соответствующий эталон строится автоматически, без вмешательства пользователя.

Отметим в заключение два момента: 1. Хотя дело заканчивается распознаванием по эталонам, применение пофонемного распознавания позволяет использовать эталоны в пределах списка пофонемно распознанных словоформ, который на порядки меньше

исходного полного словаря словоформ. 2. Программа распознавания словесного текста организована так, что позволяет сразу осуществлять голосовой набор, постепенно обучаясь в ходе работы и постепенно освобождая пользователя от дополнительного вмешательства. При этом словарь для распознавания и начальный вариант таблицы сочетаемости, конечно, должны быть заранее заданы.

# Фрагмент базы сочетаемости

Приведем в заключение фрагмент базы сочетаемости, построенной по содержанию 1-го тома книги Г.М. Фихтенгольца «Курс дифференциального и интегрального исчисления». Слева лемма слова (причастие мы трактуем как самостоятельную часть речи), справа – леммы слов, которые могут за ним следовать в данном тексте:

выраженный – в, и, обычный, так, уравнение, через, этот, явный.

выразить – и, из, кривая, он, поверхность, производная, среднее, средний, так, теперь, тот, через, число, эволюта, этот, явный.

выразиться – и, линейно, по, предел, произведение, простой, следующий, так, теперь, уравнение, формула, через.

вырезаемый – квадрат.

вырезать – по.

высказанный – в, выше, соображение, условие, утверждение.

высказать – этот.

высокий – степень, часть.

высота – использоваться, место, над, отдельный, пирамида, по, стрелка, то, треугольник, у.

# Выводы

Учет сочетаемости слов при распознавании предложения позволяет на каждом шаге, начиная со второго, во много раз сократить число кандидатов на распознавание.

# Литература

1. Шелепов В.Ю. Построение системы голосового управления компьютером на примере задачи набора математических формул / В.Ю. Шелепов, А.В. Ниценко, А.В. Жук // Искусственный интеллект. – 2010. – № 3. – С. 259-267.

***А.В. Ніценко, В.Ю. Шелєпов, Д.С. Азаренко***

**Про комп'ютерне розпізнавання мовлення з урахуванням семантики вимовленого**

Для успішного розпізнавання мовлення людиною найважливіше значення має використання смислового зв'язку слів, що розпізнаються. Більш того, успішність мовленнєвої комунікації полягає, зрештою, у досягненні збігу смислу вимовленого та сприйнятого. Стаття присвячена розробленій авторами програмі розпізнавання мовлення, яка використовує смислову сполучуваність слів у реченні. Викладення ведеться на прикладі набору математичних текстів.

***A.V. Nicenko, V.Ju. Shelepov, D.S. Azarenko***

**On Speech Recognition with a Glance of Semantics**

Key importance for successful human speech recognition is using semantic relationship of identifiable words. The article is devoted to the author's speech recognition program, which uses semantic compatibility words in sentence. Exposition for the case of mathematical text dictation.

*Статья поступила в редакцию 31.05.2011.*